CLAIMS:

1. A similar document search method of searching for a similar document similar to a specified document, comprising:

a first extracting step of extracting at least one characteristic word candidate as a candidate for a characteristic word from a seeds document including desired retrieval contents;

a second extracting step of extracting as characteristic words of the seeds document, when the characteristic word candidate extracted by said first extracting step is a compound characteristic word including a plurality of characteristic words, the compound characteristic word and constituent characteristic words included in the compound characteristic word from the characteristic word candidate;

a step of calculating, according to the characteristic words extracted by said second extracting step, similarity between the seeds document and a registration document; and

a step of outputting a retrieval result as a result of the similarity calculated by said similarity calculating step.

2. A similar document search method according to claim 1, wherein said second extracting step includes a step of determining, when constituent characteristic word information indicating constituent characteristic

words of the characteristic word is registered to the characteristic word corresponding to the characteristic word candidate extracted by said first extracting step, that the characteristic word candidate is a compound characteristic word.

3. A similar document search method according to claim 1, wherein said similarity calculating step includes:

a step of calculating a weighting coefficient corresponding to a distance between a constituent characteristic word and another constituent characteristic word which are extracted from one compound characteristic word; and

a step of calculating similarity by multiplying the weighting coefficient.

4. A similar document search system for searching for a similar document similar to a specified document, comprising:

a document analyzer processor for extracting at least one characteristic word candidate as a candidate for a characteristic word from a seeds document including desired retrieval contents;

a characteristic word extractor processor for extracting as characteristic words of the seeds document, when the characteristic word candidate extracted by said document analyzer processor is a compound characteristic word including a plurality of characteristic words, the compound characteristic word

and constituent characteristic words included in the compound characteristic word from the characteristic word candidate;

a seeds document similarity calculator processor for calculating, according to the characteristic words extracted by said characteristic word extractor processor, similarity between the seeds document and a registration document; and

a retrieval result output processor for outputting a retrieval result as a result of the similarity calculated by said seeds document similarity calculator processor.

5. A similar document search system according to claim 4, wherein said characteristic word extractor processor includes a compound characteristic word determiner processor for determining, when constituent characteristic word information indicating constituent characteristic words of the characteristic word is registered to the characteristic word corresponding to the characteristic word candidate extracted by said document analyzer processor, that the characteristic word candidate is a compound characteristic word.

6. A similar document search system according to claim 4, wherein said seeds document similarity calculator processor includes:

a weighting coefficient calculator processor for calculating a weighting coefficient corresponding to a distance between a constituent characteristic word

and another constituent characteristic word which are extracted from one compound characteristic word; and

a calculator processor for calculating similarity by multiplying the weighting coefficient.

7. A program product for making a computer operate as a similar document search system for searching for a similar document similar to a specified document, comprising:

a document analyzer processor program for extracting at least one characteristic word candidate as a candidate for a characteristic word from a seeds document including desired retrieval contents;

a characteristic word extractor processor program for extracting as characteristic words of the seeds document, when the characteristic word candidate extracted by said document analyzer processor program is a compound characteristic word including a plurality of characteristic words, the compound characteristic word and constituent characteristic words included in the compound characteristic word from the characteristic word candidate;

a seeds document similarity calculator processor program for calculating, according to the characteristic words extracted by said characteristic word extractor processor program, similarity between the seeds document and a registration document; and

a retrieval result output processor program for outputting a retrieval result as a result of the

similarity calculated by said seeds document similarity calculator processor program.

8. A program product for making a computer operate as a similar document search system according to claim 7, wherein said characteristic word extractor processor program includes a compound characteristic word determiner processor program for determining, when constituent characteristic word information indicating constituent characteristic words of the characteristic word is registered to the characteristic word corresponding to the characteristic word candidate extracted by said document analyzer processor program, that the characteristic word candidate is a compound characteristic word.

9. A program product for making a computer operate as a similar document search system according to claim 7, wherein said seeds document similarity calculator processor program includes:

a weighting coefficient calculator processor program for calculating a weighting coefficient corresponding to a distance between a constituent characteristic word and another constituent characteristic word which are extracted from one compound characteristic word; and

a calculator processor program for calculating similarity by multiplying the weighting coefficient.